Subjective and Objective Study of Different Video Coders : Application to Video Surveillance

Ludovic Quintard, Mohamed-Chaker Larabi and Christine Fernandez-Maloigne
SIC labotory FRE 2731
University of Poitiers
Blvd Marie & Pierre Curie, BP30179
86962 Futuroscope Cedex
{***Quintard,larabi,maloigne***} ***@sic.univ-poitiers.fr***

Abstract

We propose a complete study in order to give a number of recommendations for MPEG-x and MJPEG2000 coders for video surveillance application. This study aims to achieve low bitrates for the different coders but similar perceptual quality than a hardware one based on MJPEG at 5.6 Mbit/s. First, we propose a subjective test according to the ITU recommendations, the Double Stimulus Impairment Scale (DSIS) which measures the global sum of perceived degradation, as well as the Double Stimulus Continuous Quality Scale (DSCQS), which measures the visual impression for each video. Then, an objective study has been carried out. The three metrics used for this objective test are: Peak signal Noise Ratio (PSNR), the Universal Quality Index and a new metric that we have developed. Finally, a study of the correlation between the subjective and objective tests are performed in order to make some recommendations about the bitrate achieving similar quality.

Introduction

Video compression plays an important role in the framework of video surveillance. Moreover, the video needs a very big storage capacity. For this, compression could reduce this capacity to better handle the data. Many codecs are used, some of them use a frame by frame methodology, like MotionJPEG (MJPEG) and MJPEG2000 [9] when the others work with a temporal redundancy aspect as Moving Picture Experts Group (MPEG) family [10, 11, 12]. In major cases, the codec introduces some artifacts in the video [7] except in lossless compression. These artifacts affect the visual quality of the compressed video.

However, for the video surveillance applications, the codec should not give a bad visual quality because we need to recognize the face of the person for example. The main way to achieve a qualitative measurement is to proceed with a joint subjective and objective study. The subjective methods exploit the judgment of human observer and need specific conditions such as normalized room, controlled lighting, etc. When the subjective test is performed and the statistical analysis is completed, the obtained data is considered as coherent and could be recorded in the subjective report.

The objective methods are based on the impairments measurement between the original and the compressed video. These can use simple mathematical measurement like Peak Signal Noise Ratio (PSNR) or can integrate some properties of the Human Visual System (HVS). After these two tests, it is necessary to study the correlation between the subjective test and the objective one. For this, VQEG (Video Quality Expert Groups) gives some recommendations to correlate subjective and objective results [1].

This paper describes the experiments that we have conducted with different types of video coders. The purpose is to give recommendations on video compression bitrates for different codecs allowing to obtain the same visual quality than a fixed hardware coder (MJPEG at 5.36 Mbit/s) and to give some quality threshold for the used metrics for future evaluations. The tested codecs are an implementation of MJPEG2000, MPEG-1, MPEG-2 and MPEG-4. The videos have been selected with medium size Common Intermediate Format (CIF) and are representative of the video database on which we worked.

Subjective ratings were obtained for the resulting test sequences using two methodologies defined by ITU-R Recommendation BT.500 [2], namely Double Stimulus Impairment Scale (DSIS) and Double Stimulus Continuous Quality Scale (DSCQS). Objective ratings were obtained by using three measurements. The PSNR and two others that integrate some properties of the HVS namely the Universal Image Quality Index [3], and a new metric [4] that we have developed for this study.

This paper is organized as follows. In the section 2, we describe the subjective assessment methodology. The objective assessment is presented and its correlation with the subjective one is studied in section 3. Finally, we give some conclusions and future works in section 4.

Subjective Assessment
Methodology

For the test sequences, the videos used for subjective performance evaluation represent the real condition for the video surveillance. Four codecs were tested MJPEG2000, MPEG-1, MPEG-2 and MPEG-4. According to the coder, different bitrates are tested. The bitrates are chosen linearly, we divide the bitrate reference (5.6 Mbit/s) by 2, 3, 4, 5, etc.
Because the technologies of the coders, some coders are tested with a higher number of bitrate. The table 1 shows the bitrate scale for the different coders.

The viewing conditions are in conformity with the ITU Recommendations as described in [2, 5]. The Figure 1 shows an example of the laboratory setup. The distance between the screen and the observer was of 60 centimeters to respect the viewing conditions of the supervisor in real application. The monitor used

Coders bitrate

MPEG1	0.76 to 5.36 Mbit/s
MPEG2	0.67 to 5.36 Mbit/s
MPEG4	0.357 to 5.36 Mbit/s
MJPEG2000	0.76 to 5.36 Mbit/s

in the subjective assessment is a calibrated CRT display (24" designed by Sony). Each session was limited to 25 minutes in order to maintain the attention of the observer. The walls are of a neutral gray. All clips were displayed at 25fps, and all clips were shown at the same bitrate (video were first decompressed).

Assessment Methods

The two assessment methods used are specified in ITU-R recommendation BT.500 [2]. The perceptual quality needs to be the same as a MJPEG codec for a bitrate at 5.36Mbit/s. To approve the last point, we have to compare the MJPEG video with the other video coders. This is the subject of the Double Stimulus Impairment Scale (DSIS). In the other case, we want to compare the coders between them. This is the Double Stimulus Continuous Quality Scale. - In the Double Stimulus Impairment Scale (DSIS) test, observers are shown multiple sequence pairs consisting of a reference (MJPEG at 5.36Mbit/s in this case) and a test sequence (MJPEG2000, MPEG-1, MPEG-2, MPEG-4). They score the test sequence on a discrete five level scale ranging from "imperceptible" to "very annoying". - In the Double Stimulus Continuous Quality Scale (DSCQS) test, observers are shown multiple sequence pairs consisting of video "A" and video "B" where "A" and "B" could be MJPEG2000, MPEG-1, MPEG-2 or MPEG-4. They score each video on a linear scale ranging from "very good" to "very bad".

Figure 2, shows the scoring interfaces designed for this study. For each sequence pairs test a small sequence (2 seconds) announce the following video. This sequence allows to inhibit the short memory of the observer. Each video sequence lasts 12 seconds. The scoring time is unlimited.

Observers

For this study, 23 non-expert observers have participated to the test session. Each observer was tested for this visual acuity (*Snellen test*) and his color vision (*Ishihara test*). These tests are performed in order to detect color or visual blindness that could distort the results.

Figure 1. *Example of laboratory setup*

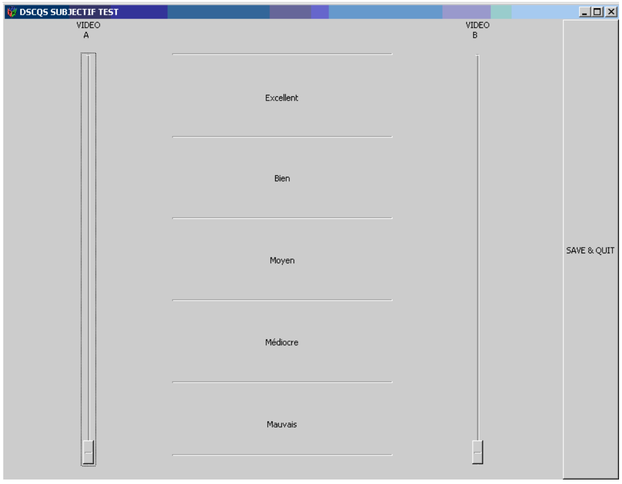

Figure 2. *Scoring Interface designed for the subjective experiments DSCQS.*

Subjective data analysis

After the subjective test, it is necessary to compute the results in order to obtain the Mean Opinion Score (MOS) and the 95% confidence interval [2]. These results are computed with the population that has passed successfully the kurtosis test [2]. This test allows to keep only the observers having a stabilized opinion. To facilitate the analysis, a digital value is given for each quality scale. For example 5 indicates the "best quality" and 1 the "worst quality", in the DSIS test.

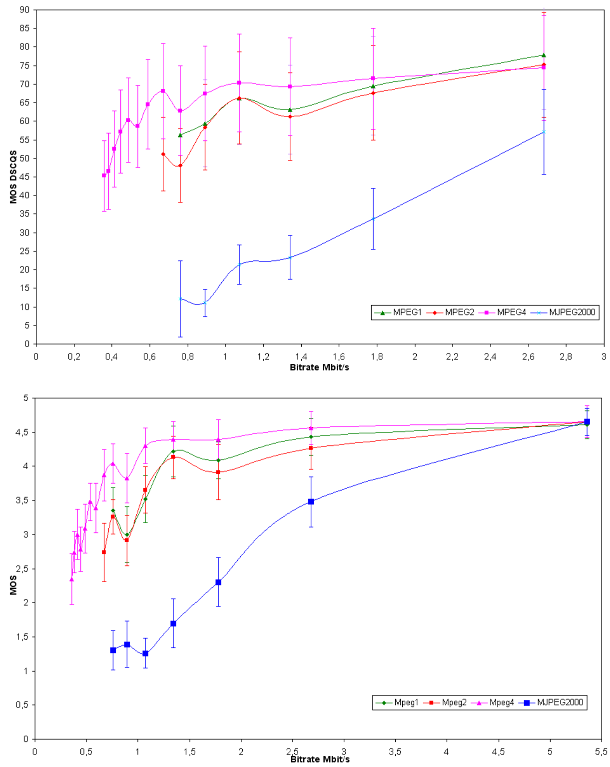

Figure 3. *MOS DSCQS (top) MOS DSIS (bottom) vs. bitrate. The vertical bar indicates the ranges of 95% confidence interval*

From figure 3, it is possible to notice that the two tests carry out toward the same conclusion; MPEG-4 coder has a better perceptual quality at low bitrates than the MPEG-1 and MPEG-2 coders. This result is due to the technology of the coders, notably the precision for the motion vectors. Moreover, the MJPEG2000 gives the worst results. These two tests are very interesting because they allow to define the perceptual quality of the video. In fact, between the scale 60 - 80 for the DSQCS test observers find the video of good quality, corresponding to a bitrate between 0.5 and 5.36 Mbit/s or 1.5 and 5.36 Mbit/s according to the coder. If we want a perceptual quality equal to the one obtained with MJPEG at 5.36Mbit/s, it is necessary to have a MOS between 3.5 and 5 for the DSIS test. These values were obtained for a bitrate equal to 1Mbit/s for the MPEG-1 and MPEG-2 codecs, 0.76 Mbit/s for the MPEG-4 codec and 3 Mbit/s for the MJPEG2000 codec.

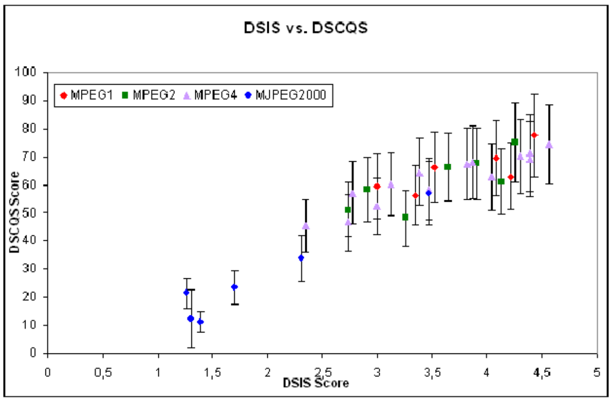

Figure 4. Comparison of DSIS and DSCQS MOS. The vertical bar indicates the ranges of DSCQS scores (95% confidence interval) within the corresponding DSIS score.

The figure 4 allows to make a comparison between the DSIS and the DSCQS tests. A linear correlation could be seen. The Pearson coefficient allows to know the correlation between two set of data [2]. Here, the Pearson correlation is about 94%. So, we are able to say that the DSIS and the DSCQS data are coherent and : "The quality of the reference video that seems good for the observer is equivalent to quality with a bitrate between 0.5-1.5 Mbit/s obtained with the coders."

Objective Assessment

Subjective assessment is expensive and time consuming. Therefore, an objective evaluation metric would be more interesting. The main idea is to determine a threshold, with regards to the subjective assessment. To attest that a metric gives good results according to the human subjective judgment it is necessary to study the correlation between the different information. For this, some tools could be used [8], in which we have the Pearson linear correlation which gives the correlation between the real MOS and the predicted one (metrics results).

Here we propose to test three different metrics. The three objective metrics need to compute rapidly for that, the three objective metrics used are the PSNR, A Universal Image Quality Index [3], and a New Metric that we have developed [4].
Peak Signal Noise Ratio (PSNR) compares the difference between pixels frames, when two frames are equal the value is about 40dB when they are different the value is about 20dB. These metric is not always correlated with the human judgment, nevertheless in the case of video compression a good correlation can be done.
The HVS is a very good contrast perception. So, the universal Image Quality Index [3] and the New Metric [4] works in the local contrast of the picture. They have respectively a range between -1 to 1 and 0 to 1. Where 1 [3] or 0 [4] is obtained when the two pictures have the same perceptual quality, -1 or 1 are obtained when the perceptual quality between two picture are worst.
The image Quality indexing approach is the following:

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2+\sigma_y^2)[(\bar{x})^2+(\bar{y})^2]} \quad (1)$$

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (2)$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2, \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2 \quad (3)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) \quad (4)$$

Equations 1,2, 3 and 4 formulate the correlation between two pictures. In the formula we have in fact three components. One components for the linear correlation between x and y. The second components measure the difference between the mean value of x and y. The third components which measure the similarity of the contrast.
This metric is apply as follow :
Starting from the top-left corner of the image, a sliding window of size $A*A$ moves pixel by pixel horizontally and vertically.

The New Metric works also on the contrast but by another approach. We compute the contrast for a window of size $A*A$ by equation 5.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (5)$$

$$R_i = \bar{x} - \bar{y} \quad (6)$$

$$R = [\frac{1}{M}\sum_{i=1} M\,|R_i|] * \theta \quad (7)$$

The difference is performed window by window between the reference image and the impaired one as shown by the equation 6. The final result is obtained by equation 7 where θ represents the maximum of admissible distorsion.

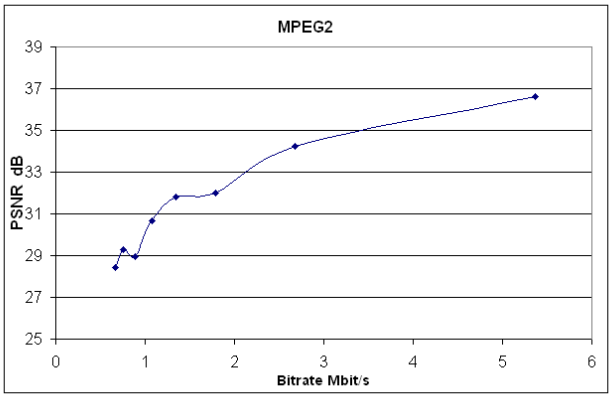

Figure 5. Example of results for MPEG2 by PSNR.

Figures 5, 6 and 7 show the results of the different metrics for the coder MPEG2. We can see the same shape for the three

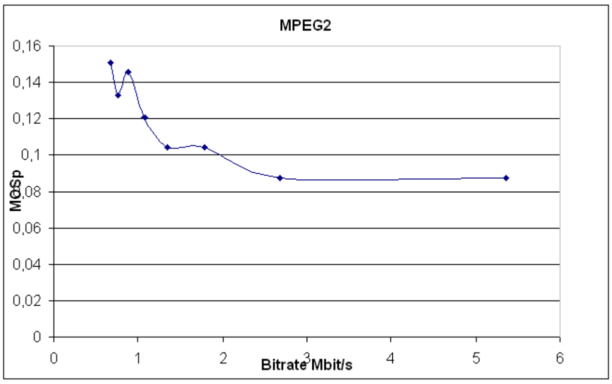

Figure 6. *Example of results for MPEG2 by the New Metric.*

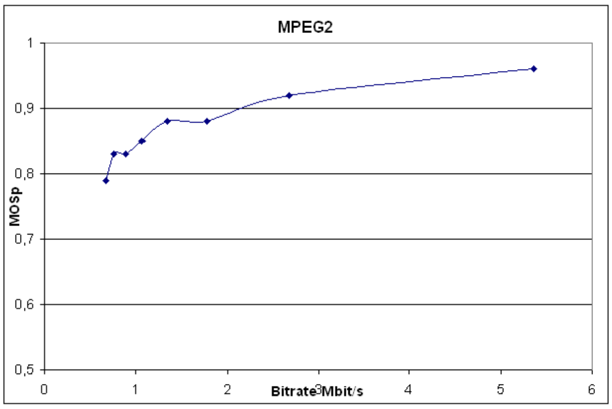

Figure 7. *Example of results for MPEG2 by the Index.*

metrics.

It is possible to find the different thresholds using the precedent figure. In fact, for each bitrate we can associate a value. This value will serve later to know if an implementation of a coder gives better or less quality.

However, before using the value or the threshold it is necessary to have a good correlation between the prediction and the value obtained by the subjective assessment. For this, some mathematical tools can be used like the *Pearson* coefficient. Figure 8, 9 and 10 show the MOS prediction *vs.* the subjective MOS. It is possible to notice that the correlation is good. In fact, when the collection of values is linear that infers a significant correlation. Indeed, the Pearson correlation values are close to 1 for all sequences.

Starting from the previous study, we can assume that we have determined the thresholds for different metrics described above which the quality will be considered as of good enough.

Conclusion

In this study we carried out some bitrate recommendations to work at same perceptual quality than a MJPEG hardware codec. Subjective and objective tests were performed. We tested four codec: MJPEG2000, MPEG-1, MPEG-2 and MPEG-4. Using the subjective test, for the same perceptual quality the bitrate recommendations is between 0.76Mbit/s for the MPEG-4 to 3Mbit/s for the MJPEG2000. The objective test give thresholds that can be use after to give the perceptual quality of a coder

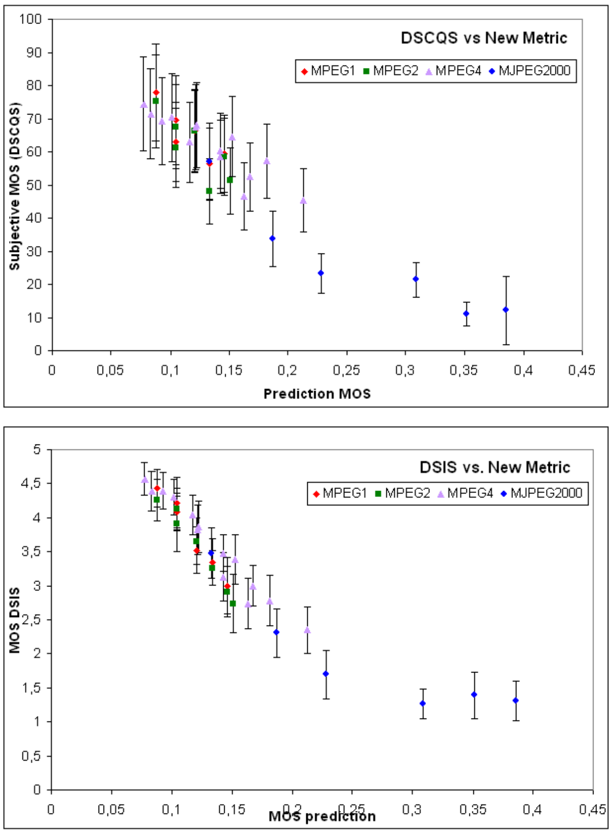

Figure 8. *Prediction MOS New Metric vs. subjective MOS. The errors bar indicates the 95% confidence interval.*

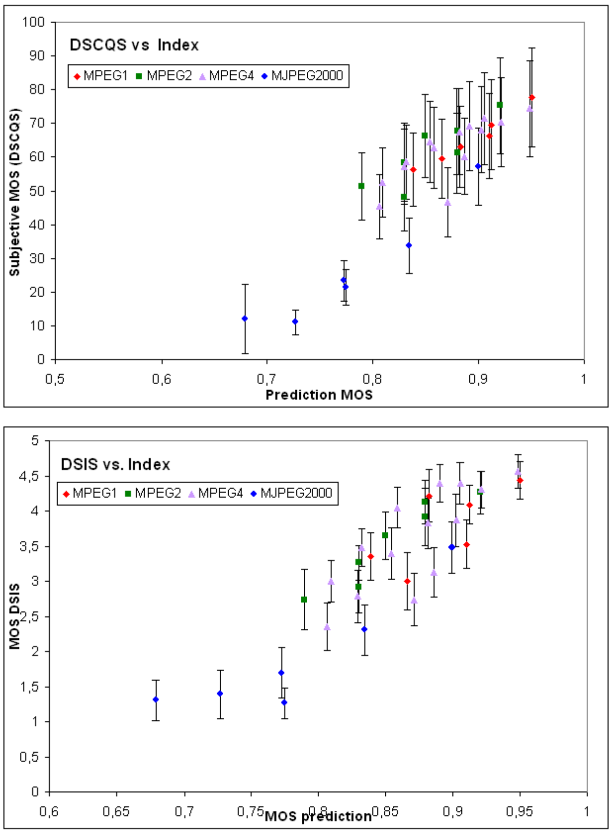

Figure 9. *Prediction MOS Index vs. subjective MOS. The errors bar indicates the 95% confidence interval.*

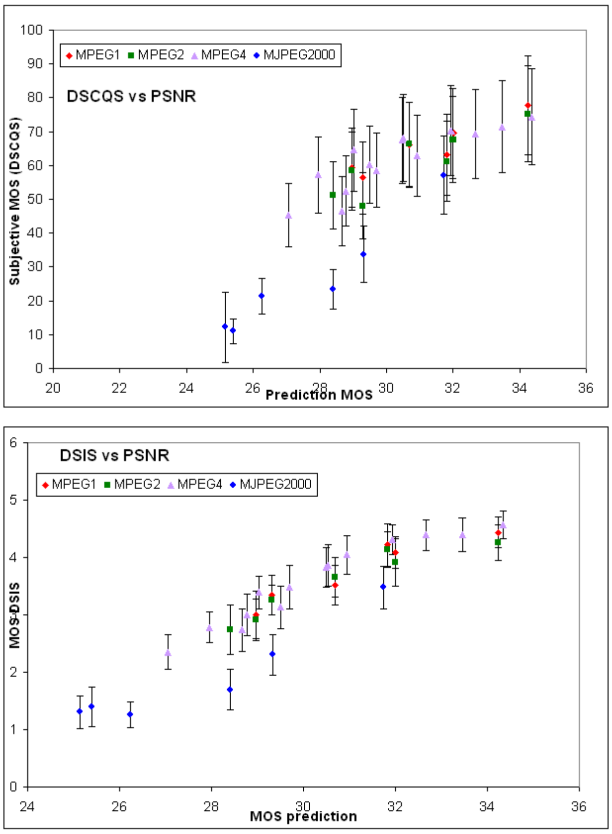

Figure 10. *Prediction MOS PSNR vs. subjective MOS. The errors bar indicates the 95% confidence interval.*

using the same data of video.

Author biography

Ludovic Quintard received his Master degree in electrical engineering from the University of Poitiers in 2005 and is currently PhD student image quality at the same university with a collaboration with the LNE (Laboratoire Nationale d'Essai). His work focuses on the development of new tools for the assessment of screen technologies.

References

[1] VQEG: Final report from the video quality expert group on the validation of objective models of video quality assessment, 2000.

[2] ITU-R Recommendation BT.500-10: Methodology for the subjective assessment of the quality of television picture, International Communication Union, 2002.

[3] Z. Wang and A. C. Bovik : A Universal Image Quality Index,IEEE Signal Processing Letters, Vol 9, n3, pp 81-84, 2002.

[4] L. Quintard and C. Larabi : Video quality : Objective an subjective assessment, technical report, university of Poitiers, 2005.

[5] ITU-R Recommendation P.910: Subjective video quality assessment methods for multimedia applications, International Communication Union,1996.

[6] S. Winkler: digital video quality vision models and metrics, John Willey & Son, New York, 2005.

[7] P. Symes, Video compression demystified, Mc Graw-Hill, New York, 2000

[8] D. Freedman, R. Pisani, and R. Purves, Statistics, W.W.Norton & Company, 3 edition, New York, 1997.

[9] MJPEG2000 ISO/IEC JPEG committee, Information technology-JPEG 2000 image coding system: Motion JPEG 2000-ISO IEC 15444-3:2002, 2002

[10] MPEG-1 ISO IEC, MPEG-1 Standard, ISO IEC, Geneva,Switzerland, 1993.

[11] MPEG-2 ISO IEC, MPEG-2 Standard, Nov., Geneva,Switzerland, 1993.

[12] MPEG-4 ISO IEC MPEG-98, MPEG-4 Overview, Requirements Group, Doc. N2459, Geneva,Switzerland, Oct. 1998.